The Report Committee for Xiaolei Xue
Certifies that this is the approved version of the following report:

A Logistic Regression Analysis for Potentially Insolvent Status of Life Insurers in the United States

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ________________________________
Thomas W. Sager

____
Margaret E. Myers

A Logistic Regression Analysis for Potentially Insolvent Status of Life Insurers in the United States

by

Xiaolei Xue, B.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2011

Dedication

This Thesis is dedicated to my husband, son and parents who provided support with all their efforts.

Acknowledgements

Special thanks to the following persons who helped me in my graduate career through their own unique ways.

To my advisor, Dr. Thomas Sager, for his generous helps on design, discussion, analysis and providing NAIC database.

To my committee, Dr. Margaret Myers, for her precious suggestions, time and patience.

And to my family for their support.

May 04, 2011

Abstract

A Logistic Regression Analysis for Potentially Insolvent Status of Life Insurers in the United States

Xiaolei Xue, M.S.STAT.

The University of Texas at Austin, 2011

Supervisor: Thomas W. Sager

This study focused on identifying factors that significantly affect the potentially insolvent status of life insurers. The potentially insolvent status is indicated based on insurer's Risk-based capital ratio (RBC ratio) reported in the National Association of Insurance Commissioners (NAIC) database of life insurers' annual statements. A logistic regression analysis is performed to explore the relationship between the RBC insolvent indicator and a set of explanatory variables including insurer's size, capital, governance structure, membership in a group of affiliated companies, and various risk measures during the 2006-2008 period.

The results suggest that the probability of potential insolvency for an individual insurer is significantly affected by its size, capital-to-asset ratio, returns on capital, health product risk and proportion of products reinsured. It could be also possibly affected by

the insurer's regulatory asset risk. However, the results indicate that the probability is not significant related to the insurer's annuity product risk, opportunity asset risk, governance structure and its membership in a group of affiliated companies. On average, by holding all other explanatory variables constant, every 1% increase in total assets will result in a decrease of 0.19 to 0.36% on the odds of potentially insolvent rates; every 0.01 unit increase in capital-to-asset ratio will result in a decrease of a multiplicative factor of 0.951 to 0.956 on the odds; every 0.01 unit increase in return on capital will result in a decrease of a multiplicative factor of 0.984 to 0.985 on the odds; every 0.01 unit increase in health product risk will result in an increase of a multiplicative factor of 1.021 to 1.031 on the odds; and every 0.01 unit increase in proportion of products reinsured will result in an increase of a multiplicative factor of 1.015 to 1.026 on the odds.

The assumptions of independency and absence of harmful multicolliearity are both valid for this logistic model, suggesting that the model is adequate and the conclusion is warranted. Although the potentially insolvent indicator, instead of the real insolvent indicator is used, this model could still be useful to identify the significant factors which affect life insurers' potentially insolvent status.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Risk-based capital ratio

The insolvency of an insurance company will not only affect its policyholders by terminating their policies, but will influence brokers, agents, reinsurer and state guaranty funds as well. Shortly after the high insolvent rate ran into the American life / health insurance industry between 1989 and 1991, the National Association of Insurance Commissioners (NAIC) established risk-based capital (RBC) standards for insurers and adopted different RBC formulas for life and health insurers in 1992 (Browne, Carson and Hyot, 1999). The standards became effective in 1994. These formulas are used to compute a minimum capital requirement for specific insurers according to their individual risk portfolios. In practice, the levels of regulatory actions are often determined by the risk-based capital ratio (RBC ratio). The RBC ratio is defined as 200 times the ratio of an insurer's total adjusted capital (difference between assets and liabilities) to its authorized control level capital (calculated minimum required capital) (NAIC, 2009).

Different kinds of regulatory actions are required when insurers fail to meet certain minimum RBC thresholds. In brief, no action is required for RBC ratio greater than 200. If an insurer has an RBC ratio of 200, then its available capital exactly matches the level required to cover its estimated asset risk exposure. If the RBC ratio is below 200, then its available capital is inadequate to cover its asset risk. Company level actions, including comprehensive financial plans and report submission to a regulator, are required when the RBC ratio is between 150 and 200. Regulatory level actions, including necessary exams on insurer's financial position and appropriate financial orders are required when the RBC ratio is between 100 and 150. For an RBC ratio below 100, the regulator has the option of taking control of the insurer if it is above 70 (authorized control level), and if it is below 70, it is required to place the insurer under control (mandatory control level) (NAIC, 2009).

After the economic crisis starting in 2008, especially after American International Group, Inc. (AIG), the largest insurance company in the world, suffered a liquidity crisis that ended with a federal government bailout, the NAIC tightened its regulatory policies and introduced additional trend tests. Insurers with an RBC ratio from 200 to 250 in the life insurance industry or 200 to 300 in the health insurance industry will trigger the trend test, and a negative trend below a certain level will result in a company level action (NAIC, 2009).

Research on the RBC ratio as a regulatory tool in the insurance industry has been conducted during the past decade. In the property and casualty area, Cummins, Harrington and Klein (1995) concluded that RBC ratios were significantly different for insolvent and surviving companies, and the prediction RBC model was more successful in predicting smaller firm insolvencies. Various comparisons on predictive capabilities were also conducted between the RBC ratio and other solvency monitoring tools, including the NAIC Financial Analysis Solvency Tools (FAST) score, AM Best's Capital Adequacy (BCA) and a cash flow simulation model (Grace, Harrington and Klein, 1998; Pottier and Sommer, 2002; Cummins, Grace and Phillips, 1997). The life and health area has relatively few studies when compared to the property and casualty industry. Pottier and Sommer (1997) compared the RBC ratio to other insurance organization ratings, Ryan and Schellhorn (2000) examined the impact of the RBC regulations on life insurer's efficiency, and Baranoff and Sager (2002) did an exploratory work in simultaneous interrelation analysis among capital, assets and product risk.

An indicator representing the actual insurer insolvent status is optimal as response variable to identify the factors which significantly affect insurer's insolvency. However, there was relatively few number of insolvencies in any given year recently. That is, there might not be enough data to draw statistically meaningful inference when using actual insolvent indicator. Notice that it is very likely that actual insolvencies will come from the low RBC ratio category, we adopted RBC ratio as a proxy indicator for insurer insolvent status. RBC ratio was selected not only because it was generated as a regulation tool updated annually for all insurers, but was reported as a good predictor of insurer insolvency as well (Pottier and Sommer, 1997; Ryan and Schellhorn, 2000). While previous studies were focused on using RBC ratio directly to predict the insurer's insolvency, this study investigate the relationship between a proxy for the potential for insolvency and a set of factors including insurer's size, structure, and risks. RBC ratio is no longer used as a predictor but used as an indicator of potential for insolvency. The goal of this study is to identify factors that significantly affect an individual life-health insurer's solvency. The RBC ratio will be converted to a binary variable using an empirical cut-off value based on the distribution of the RBC ratio to distinguish between "potentially insolvent" and "surviving" status. A logistic regression will then be performed to examine the relationship between the RBC indicator (response variable) and a series of explanatory variables, including but not limited to insurer's size, insurer's governance structure, retained earnings, asset risks, and product risks. The results will be interpreted to better understand the prediction of life-health insurers' insolvency and survival.

1.2 Summary of chapters

Chapter 1 provides an introduction to the motivations for this study as well as the background information on RBC ratios.

Chapter 2 focuses on the methodologies used in the study, including introduction to logistic regression, data processing before feeding into logistic regression and defining variables.

Chapter 3 illustrates results and draws conclusions, including analysis, outputs interpretation, model comparisons and assumption validations.

1.3 References

Baranoff, E. G. and Sager,T. W., 2002, The relations among asset risk, product risk, and capital in the life insurance industry, Journal of Banking & Finance 26 (2002) 1181–1197.

Browne, M. J., Carson, J. M., Hoyt, R. E., 1999, Economic and Market Predictors of insolvencies in the Life-Health Insurance Industry. Journal of Risk and Insurance, Vol. 66, No. 4, pp. 643-659.

Cummins, J. D., Grace, M. F., and Phillips, R. D., 1997, "Regulatory Solvency Prediction in Property-Liability Insurance: Risk-Based Capital, Audit Ratios, and Cash Flow Simulation." Center for Risk Management and Insurance Research Working Paper 97-4, Georgia State University.

Cummins, J. D., Scott E. H., and Robert W. K., 1995, “Insolvency Experience, Risk-Based Capital, and Prompt Corrective Action in Property-Liability Insurance.” Journal of Banking and Finance, 19: 511-527.

Grace, M. F., Scott E. H., and Robert W. K., 1998, “Risk-Based Capital and Solvency Screening in Property-Liability Insurance: Hypotheses and Empirical Tests.” Journal of Risk and Insurance, 65: 213-243.

NAIC, 2009, NAIC Capital Adequacy Task Force - Risk Based Capital Overview. www.naic.org/documents/committees_e_capad_RBCoverview.pdf

Pottier, S.W., Sommer, D.W., 1997. Life insurer risk-based capital measures: Does the source matter? Journal of Insurance Regulation 16 (2), 179–196.

Pottier, S.W., and Sommer, D.W., 2002, The Effectiveness of Public and Private Sector Summary Risk Measures in Predicting Insurer Insolvencies, Journal of Financial Services Research, 21(1): 101–116.

Ryan Jr., H.E., Schellhorn, C.D., 2000. Life insurer cost efficiency before and after implementation of the NAIC risk-based capital standards. Journal of Insurance Regulation 18 (3), 362–384.

Chapter 2: Methodology

In order to obtain the probability of the insolvency of an individual insurer with the given situations about capital, asset and risks, a logistic regression analysis will be conducted by regressing the RBC indicator variable on a series of explanatory variables relating to capital, assets, returns, and risks. Significant regressors will be identified and corresponding coefficients will be used to predict the probabilities.

2.1 LOGISTIC REGRESSION ANALYSIS

Logistic regression is commonly used to predict the probability of event occurrence. The prediction is achieved by fitting the data to an “S-shaped” logistic curve (Cox and Snell, 1989).

2.1.1 Logistic function

As the basis of logistic regression analysis, logistic function defines an “S-shaped” curve by the following equation:

$$p(y|\boldsymbol{X}) = \frac{\boldsymbol{e}^{X\beta}}{1 + e^{X\beta}} = \frac{1}{1 + e^{-X\beta}} \quad (1)$$

where y is a binary indicator with values 0 and 1, and p(y|X) is the probability of “success” (i.e., y=1), given a set of explanatory variables X, and β is a vector of coefficients. For any combination of X, the corresponding output p(y|X) is confined to values between 0 and 1.The graph representing the logistic function is shown as Figure 1 for a univariate X.

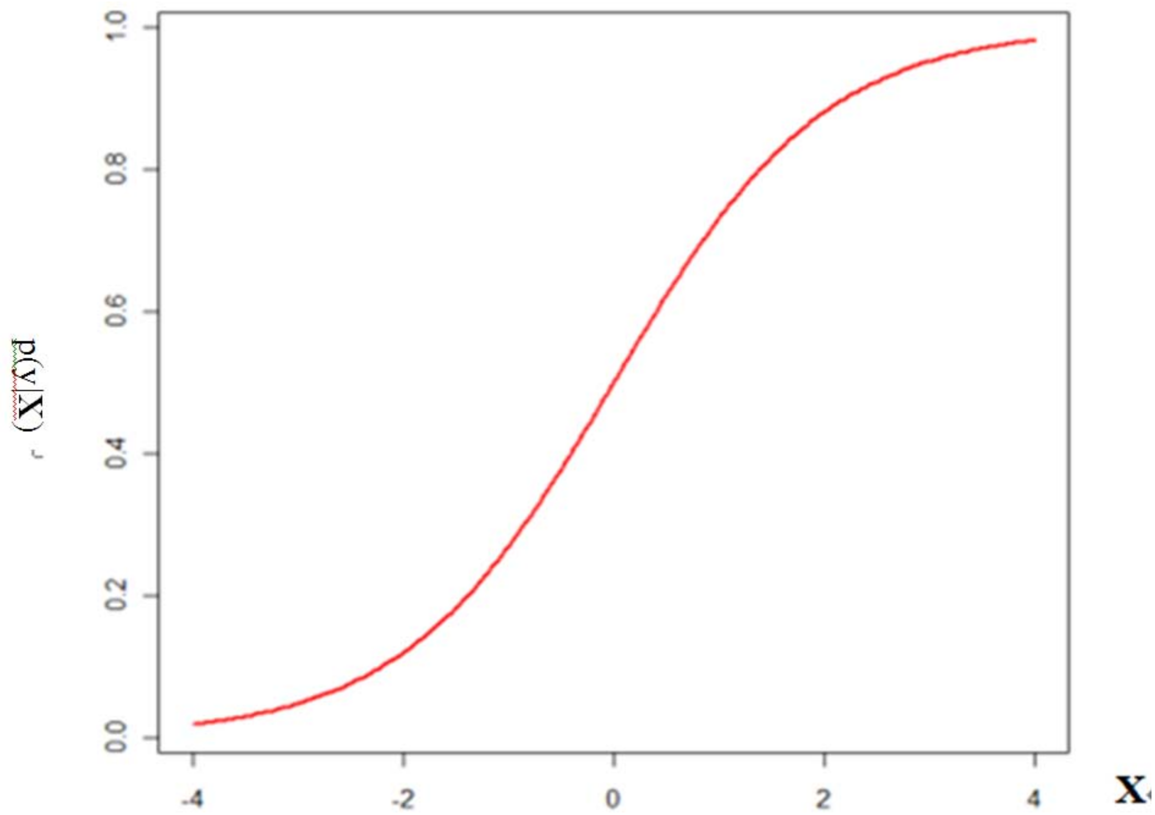

Figure 1: Logistic function. p(y|X) represents the probability of "success."

2.1.2 Logistic regression

Logistic regression is a nonlinear regression and requires the response variable to be a binary or dichotomous variable. A binary variable is a variable with two categories of outcomes, such as success (1) or failure (0). If each trial follows an identical Bernoulli distribution, the number of success (S) out of N independent trials (with the same probability of success as p) will have a binomial distribution (noted as S~ Binomial (N, p)).

The logistic density function reveals the probability of "success," which is affected by a set of explanatory variables as expressed in Equation (1).

$$p(y|\boldsymbol{X}) = \frac{1}{1 + e^{-X\beta}}$$

where $y = \boldsymbol{X}\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$. The X's are explanatory variables, such as insurer's size, insurer's governance structure, retained earnings, asset risks, and product risks. β_0 is defined as the "intercept." Mathematically, it is equal to the value of y when all explanatory variables Xs are set to be zeros. In practice, the interpretation of β_0 could have no meaningful explanations for certain specific problems. $\beta_1, \beta_2, \ldots \beta_m$ are defined as the "regression coefficients" of explanatory variables X_1, $X_2 \ldots X_m$ respectively. The variability of the response variable is affected by explanatory variables in the weight of contributions described by corresponding regression coefficients. In brief, a positive (negative) regression coefficient suggests an increase (decrease) in probability of "success" as the associated explanatory variable increases, with influence levels indicated by the magnitudes of the coefficients.

By rearranging Equation (1) and taking the logarithm on both sides, we get the following logistic regression model:

$$\mathrm{logit(p)} = \ln\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \qquad (2)$$

The ratio of the probability of success to the probability of failure is defined as the *odds of success*, and the logarithm of the odds is known as a *logit*.

More practically, the probability of success can be predicted by

$$\hat{p} = \frac{1}{1+e^{-(\hat{\beta}X)}} \tag{3}$$

where $\hat{p}$ is the predicted probability of success and $\hat{\beta}$ is the vector of regression coefficients (Hosmer and Lemeshow, 2000; Hilbe, 2009 and Agresti, 2002).

The odds-ratio is the extended version of odds and describes the odds of success associated with one group compared to another (Powers and Xie, 1999). For instance, consider two binary predictors X_1 and X_2, each with possible values 0 and 1. X_1 could be governance type (1=stock, 0=mutual); X_2 could be membership in an affiliated group of insurers (1=yes, 0=no). Now consider the group of unaffiliated stock insurers and the group of affiliated mutual insurers. If the probabilities of success of two groups are p1 and p2 respectively, the corresponding odds will be $\frac{p_1}{1-p_1}$ and $\frac{p_2}{1-p_2}$, and the corresponding logits for a two-predictor model will be:

$$logit(p_1) = \beta_0 + \beta_1$$

$$logit(p_2) = \beta_0 + \beta_2$$

The odds-ratio of group1 versus group2 is expressed below:

$$odds\ ratio = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0+\beta_2}} = e^{(\beta_1-\beta_2)} \tag{4}$$

2.1.3 Interpreting the regression coefficients

The regression coefficients are commonly estimated using the maximum likelihood method. The interpretation of regression coefficients in logistic regression is different from that of ordinary linear regression, but is related to that in multiple linear regression with response variables in logarithm form. The interpretation can be further separated into two forms of the relationship:

- Logit(p) to X
- Logit(p) to logX

For the form of Logit(p) to X, the coefficient is interpreted as the expected change in the logit (natural log of odds) for unit increase in explanatory variable X_1 $(X_2, \ldots, X_m)$, holding all other variables constant. A more meaningful interpretation is often arrived at by transforming the coefficient exponentially. For example, instead of interpreting $\widehat{\beta_1}$ as the expected change in logit, $e^{\widehat{\beta_1}}$ is explained as the multiplicative change in the odds of success per unit increase in the explanatory variable X_1, holding all other variables constant. In other words, the odds will increase by $100 \times (e^{\widehat{\beta_1}} - 1)\%$ for unit increase in the explanatory variableX_1, holding all other variables constant.

For category of Logit(p) to logX, the interpretation is based on the following fact. Assuming there is a 1% increase in X (i.e. X'=1.01X) and holding all other variables constant, the expected increase in logit is then

ln(odds')-ln(odds)=βlnX'-βlnX;

i.e. $\ln(\frac{odds'}{odds})=\beta\ln(\frac{X'}{X})=\ln(1.01)^{\beta}$ or

$$\frac{odds'}{odds}=(1.01)^{\beta} \qquad (5)$$

The right side of Equation (5) can be expanded in the form of a binomial series, which is approximately 1+0.01β+R(β), where R(β) represents the remainder term of expansion and is close to 0 since it is related to the powers of 0.01. More practically, the regression coefficient can be interpreted as the following: a 1% increase in X while holding all other variables constant will result in an approximate β% increase in odds, on average.

2.2 DATA PROCESSING

The NAIC life-health insurers RBC data set was provided by Dr. Sager at the University of Texas at Austin. There were a total of 18056 records associated with the corresponding RBC ratios. The RBC ratio was defined by the following expression:

$$\text{RBC ratio}=\frac{Market\ Capital}{Authorized\ Capital}\times 200$$

The ratios were later converted into an RBC indicator variable using a cut-off value so that it could meet the requirement of a response variable of a logistic regression analysis. The cut-off value was chosen based on a preliminary study of the distribution of the RBC ratio. For demonstration purposes, a cut-off value of 200 was chosen to distinguish between the "potential insolvent insurers" (0s) and "surviving insurers" (1s). This resulted in 11.41% of a total of 18,056 records to be "potential insolvent insurers."

There were 33 records associated with a negative RBC ratio. The negative RBC ratio occurred because of accounting anomalies. Additionally, extremely large RBC ratios (greater than 10,000) usually occurred because of very small denominators (authorized capital), which could be accounted for by very small companies or companies winding up their affairs and preparing to go out of business. Therefore, all records that had the RBC ratios that were negative or greater than 10,000 were dropped. The dropped records accounted for less than 2% of all records.

2.3 EXPLANATORY VARIABLES

The following variables related to insurers' size, structure and risks were included as explanatory variables to identify the significant variables as suggested by previous researches (Baranoff and Sager, 2002; Pottier and Sommer, 1997; Shrieves and Dahl, 1992; Grace and Timme, 1992; Baranoff, 2004).

- Insurer's size measured by the logarithm of the company's total assets (LogATotal);
- A ratio of market capital to total assets (CAP);
- A return on capital (RetOnCap) as retained earnings, defined as a ratio of income to market capital;
- Product risks, including annuity risks (ProdArisk), which was defined as the ratio of annuity writings to total writings, and health risks (ProdHrisk), which was defined as the ratio of health writings to total writings. More weight was expected on product risk associated with health writings;
- proportion reinsurance writings to total writings (preinsur);
- Regulatory asset risk and opportunity asset risk in logarithm form (logpregarisk and logpopparisk);

- An indicator variable for the governance structure (Ntype: stock=1 and mutual=0);
- An indicator variable for whether or not the insurer is a member of a group of affiliated companies (Ngroup, member=1 and non-member=0);

2.4 DATA ANALYSIS

To avoid the autocorrelation between RBC indicators of successive years, records are separately analyzed year by year. For demonstration purpose, the results of the latest three years (2006, 2007, and 2008) are reported and analyzed. Descriptive statistics are summarized to show the mean, standard deviation and median of each variable. To further understanding of the relationship, a logistic regression is performed to regress the RBC indicator variable on explanatory variables listed above.

The statistical software package SAS 9.2 is used to obtain descriptive statistics, perform the logistic regression and analyze the relationship between the response variable and the explanatory variables. Predicted probabilities are plotted against different explanatory variables to visualize the relationships using statistics software package R 2.12.1.

In summary, the logistic regression is well suited for the analysis of the relationship between the binary response variable, the RBC indicator and a set of explanatory variables. The regression coefficients, when interpreted properly, not only show their relative strength of influence on the response variable, but help to predict the probability of the occurrence of an event as well. The model built using logistic regression will help to predict the potentially insolvent ratio of an individual insurer.

2.5 References

Agresti, A. 2002. Categorical Data Analysis. New York: Wiley-Interscience.

Baranoff, E. G., 2004, Risk Management and Insurance. John Wiley & Sons, Inc.

Baranoff, E. G. and Sager,T. W., 2002, The relations among asset risk, product risk, and capital in the life insurance industry, Journal of Banking & Finance 26 (2002) 1181–1197.

Cox, D. R., and Snell, E. J. 1989. Analysis of Binary Data. Second edition. London: Chapman & Hall.

Grace, M.F., Timme, S.G., 1992. An examination of cost economics in the US life insurance industry. Journal of Risk and Insurance 59, 72–103.

Hilbe, J. M. 2009. Logistic Regression Models. Chapman & Hall/CRC Press

Hosmer, D. W., and Lemeshow, S. 2000. Applied Logistic Regression. 2nd ed. John Wiley & Sons Inc.

Pottier, S.W., Sommer, D.W., 1997. Life insurer risk-based capital measures: Does the source matter? Journal of Insurance Regulation 16 (2), 179–196.

Powers, A. P., and Xie, Y. 1999. Statistical Methods for Categorical Data Analysis. Academic Press, Inc.

Shrieves, R.E., Dahl, D., 1992. The relationship between risk and capital in commercial banks. Journal of Banking and Finance 16, 439–457.

Chapter 3: Results and Conclusions

3.1 Descriptive statistics

To reveal the relationship between the indicator of "potentially insolvent" and a set of explanatory variables for individual insurers, analysis is performed year by year over the three-year period (2006-2008). There are 770 to 849 life insurers with complete records for each of the three years used in the study. The following table shows a summary of statistics for those insurers (Table 1).

Table 1. Summary statistics: 2006-2008

Variable	Label	Mean			Std. Dev.		
		2006	2007	2008	2006	2007	2008
rbc200	Indicator for insolvency(0) or survival(1)	0.923	0.933	0.896	0.266	0.249	0.305
LogATotal	log(Total assets)	19.146	19.231	19.186	2.785	2.776	2.767
CAP	Market capital / Total assets	0.309	0.312	0.314	0.268	0.270	0.283
RetOnCap	Income / Market capital	0.079	-0.015	-0.076	0.508	2.107	0.570
ProdARisk	Annuity writings / Total writings	0.171	0.168	0.180	0.302	0.301	0.311
ProdHRisk	Health writings / Total writings	0.289	0.296	0.291	0.372	0.375	0.374
Preinsur	Reinsurance writings / Total writings	0.147	0.146	0.149	0.344	0.311	0.303
logpregarisk	log(regulatory asset risk)	-4.358	-4.380	-4.390	1.135	1.132	1.087
logpopparisk	log(opportunity asset risk)	-6.089	-5.830	-5.467	0.419	0.680	0.687
Ngroup	Indicator for member of group (1=yes)	0.774	0.773	0.771	0.419	0.419	0.421
Ntype	Indicator for stock(1) or Mutual(0)	0.920	0.923	0.927	0.272	0.266	0.261

According to the table, the potentially insolvent rates are between 7% and 10% within the three-year period. The sizes of insurers are measured using a logarithm of the total assets because of the large variability reported in previous literature (Baranoff and Sager, 2002). The average of log(assets) is around 19.2. The average capital to asset ratio is about 31% for the period. The retained earnings, measured by income per capital, is the only variable that varied largely within the study period, from earning 7.9 cents per dollar in 2006 to losing 7.6 cents per dollar in 2008. Regarding the components of risk measures, annuity writings contribute around 17-18% of the average insurer's writing for the study period while health writings contribute around 29-30% of the average writings. On average, around 15% of total writing are reinsurance writing, which is the insurance of other insurance companies. Regulatory asset risk and opportunity asset risk are both reported in logarithm forms. The corresponding geometric averages are 1.3% and 0.2-0.4% for regulatory asset risk and opportunity asset risk, respectively. The means of indicators for affiliated company membership and governance type reveal that around 77% of the insurers in the database are members of affiliated groups and 92% are stock companies.

3.2 Logistic model estimation

Logistic regression analysis is performed to examine the significance of the relationship between "potentially insolvent" and previous defined explanatory variables. The proposed logistic model is as below:

$$\mathrm{Logit(p)} = \ln\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right) = \beta_0 + \beta_1 \times LogATotal + \beta_2 \times \mathrm{CAP} + \beta_3 \times \mathrm{RetonCap} + \beta_4 \times \mathrm{ProdARisk} + \beta_5 \times \mathrm{ProdHRisk} + \beta_6 \times \mathrm{Preinsur} + \beta_7 \times \mathrm{LogPRegARisk} + \beta_8 \times \mathrm{LogPOppARisk} + \beta_9 \times I_{group} + \beta_{10} \times I_{type} \quad (6)$$

Results for logistic regression analysis of the three years are shown in Table 2. Both regression coefficients and odds estimates are reported for easy interpretation.

Table 2. Model estimates for logistic model

	Analysis of Maximum Likelihood Estimates					Odds Ratio Estimates		
Year	Variable	Estimate	S.E.	Wald χ^2	p-value	Point Estimate	95% C.I.	
2006	Intercept	6.4297	2.977	4.6648	0.0308			
	LogATotal	-0.3634	0.0798	20.7681	<.0001	0.695	0.595	0.813
	CAP	-5.0391	0.9205	29.9648	<.0001	0.006	0.001	0.039
	RetOnCap	-0.3333	0.2162	2.3763	0.1232	0.717	0.469	1.095
	ProdARisk	0.489	0.8271	0.3495	0.5544	1.631	0.322	8.25
	ProdHRisk	2.886	0.4878	35.0081	<.0001	17.922	6.889	46.619
	Preinsur	1.6194	0.5861	7.6334	0.0057	5.05	1.601	15.93
	logpregarisk	0.3844	0.1475	6.7917	0.0092	1.469	1.1	1.961
	logpopparisk	0.0685	0.4489	0.0233	0.8787	1.071	0.444	2.581
	Ngroup	-0.198	0.3348	0.3496	0.5544	0.82	0.426	1.581
	Ntype	0.00257	0.5336	0	0.9962	1.003	0.352	2.853
	Pseudo R-squared=0.1024							
2007	Intercept	6.593	2.3278	8.0217	0.0046			
	LogATotal	-0.2942	0.0938	9.8428	0.0017	0.745	0.62	0.895
	CAP	-4.9819	1.0867	21.0175	<.0001	0.007	<0.001	0.058
	RetOnCap	-1.4763	0.4228	12.1901	0.0005	0.228	0.1	0.523
	ProdARisk	0.1986	0.8556	0.0539	0.8164	1.22	0.228	6.525
	ProdHRisk	2.1139	0.5237	16.2934	<.0001	8.28	2.967	23.111
	Preinsur	1.4541	0.5816	6.252	0.0124	4.281	1.369	13.383
	logpregarisk	0.1304	0.1814	0.517	0.4721	1.139	0.798	1.626
	logpopparisk	0.458	0.3516	1.6965	0.1927	1.581	0.794	3.149
	Ngroup	-0.5342	0.3594	2.2086	0.1372	0.586	0.29	1.186
	Ntype	0.1779	0.6466	0.0757	0.7832	1.195	0.336	4.243
	Pseudo R-squared=0.0951							
2008	Intercept	3.6974	2.0256	3.3319	0.0679			
	LogATotal	-0.1939	0.0782	6.1406	0.0132	0.824	0.707	0.96
	CAP	-4.5289	0.9279	23.8239	<.0001	0.011	0.002	0.067
	RetOnCap	-1.6478	0.28	34.6339	<.0001	0.192	0.111	0.333
	ProdARisk	0.4545	0.7293	0.3884	0.5331	1.575	0.377	6.579
	ProdHRisk	3.0223	0.5264	32.9631	<.0001	20.538	7.319	57.628
	Preinsur	2.533	0.552	21.056	<.0001	12.591	4.268	37.148
	logpregarisk	0.3294	0.1878	3.0749	0.0795	1.39	0.962	2.009
	logpopparisk	0.1686	0.3098	0.2962	0.5863	1.184	0.645	2.172
	Ngroup	-0.4791	0.3351	2.0437	0.1528	0.619	0.321	1.195
	Ntype	-0.2286	0.4955	0.2129	0.6445	0.796	0.301	2.101
	Pseudo R-squared=0.1548							

For all three years, the p-value for the global null hypothesis that all regression coefficients are zeros is nearly zero, suggesting that the model as a whole is significant. SAS reported that the pseudo R-squared are 0.1024, 0.0951 and 0.1548 for the years 2006, 2007 and 2008, respectively. Although the pseudo R-squared statistic in the logistic regression is not calculated in the same way as R-squared in linear regression, researchers used simulations to predict a continuous, latent variable through the ordinary least square (OLS) regression and its observed binary variable through logistic regression. Then they compare the pseudo R-squared to the OLS R-squared. The comparison result shows that pseudo-R-square sometimes tends to underestimate the proportion of variance explained (Freese and Long, 2006; Long, 1997). Therefore, the above pseudo R-squared suggest that the variability of potentially insolvent probability can be explained by the model by 10.24%, 9.51% and 15.48% for the three years.

Most explanatory variables that are identified as significant predictors are similar throughout the three years. However, some variables, such as return on capital and logarithm of proportion of regulatory asset risk, vary year by year. In detail:

- *Membership and structure*

P-values for membership of affiliated companies and governance structure are much greater than 0.05 for all three years, indicating they are not significant predictors for potentially insolvent status for individual insurers. Therefore, the expected probability of potentially insolvent status of an insurer is not significantly affected by either its membership status of affiliated companies or its governance structure as a stock / mutual company.

- *Assets and capital*

Both size (logarithm of total assets) and capital-to-asset ratio (CAP) are recognized as significant predictors in all three years. The estimates of regression coefficient for Log(Total Assets) are between -0.3634 and -0.1939, and the corresponding odds estimates are between 0.695 to 0.824. The result can be interpreted as follows: for each unit increase in the logarithm of total assets, the odds of potentially insolvent rates are expected to decrease by a multiplicative factor of 0.695 to 0.824 on average, holding all other explanatory variables constant during the 2006-2008 period. In other words, for every 1% increase in total assets, the odds of potentially insolvent rates are expected to decrease by 0.19 to 0.36% on average, holding all other explanatory variables constant during the 2006-2008 period.

The estimates of the regression coefficient for capital-to-asset ratio are between -5.0391 and -0.4529, and the corresponding odds estimates are between 0.006 and 0.011. The result can be interpreted as follows: for each 0.01 unit increase in capital-to-asset ratio, the logits of potentially insolvent rates are expected to decrease by 0.0504 to 0.0453 units on average, or the odds of potentially insolvent rates are expected to decrease by a multiplicative factor of 0.951 to 0.956, holding all other explanatory variables constant during the 2006-2008 period.

The other explanatory variable return on capital is significantly related to the potentially insolvent rate only in years 2007 and 2008. The estimates of the regression coefficient are -1.4763 and -1.6478, and the corresponding estimates for odds are 0.228 and 0.192 for year 2007 and 2008, respectively. The result can be interpreted as follows:

for each 0.01 unit increase in return on capital, the logits of potentially insolvent rates are expected to decrease by -0.0165 to -0.0148 units on average, or the odds of potentially insolvent rates are expected to decrease by a multiplicative factor of 0.984 to 0.985, holding all other explanatory variables constant for years 2007 and 2008 respectively.

The predicted probabilities of potential insolvency versus log(total assets) and capital-to-asset ratio are plotted for three years as shown in Figure 1. Other explanatory variables are held constant at mean values for continuous variables or at 1 for indicator dummy variables (i.e. member of affiliated group and stock company).

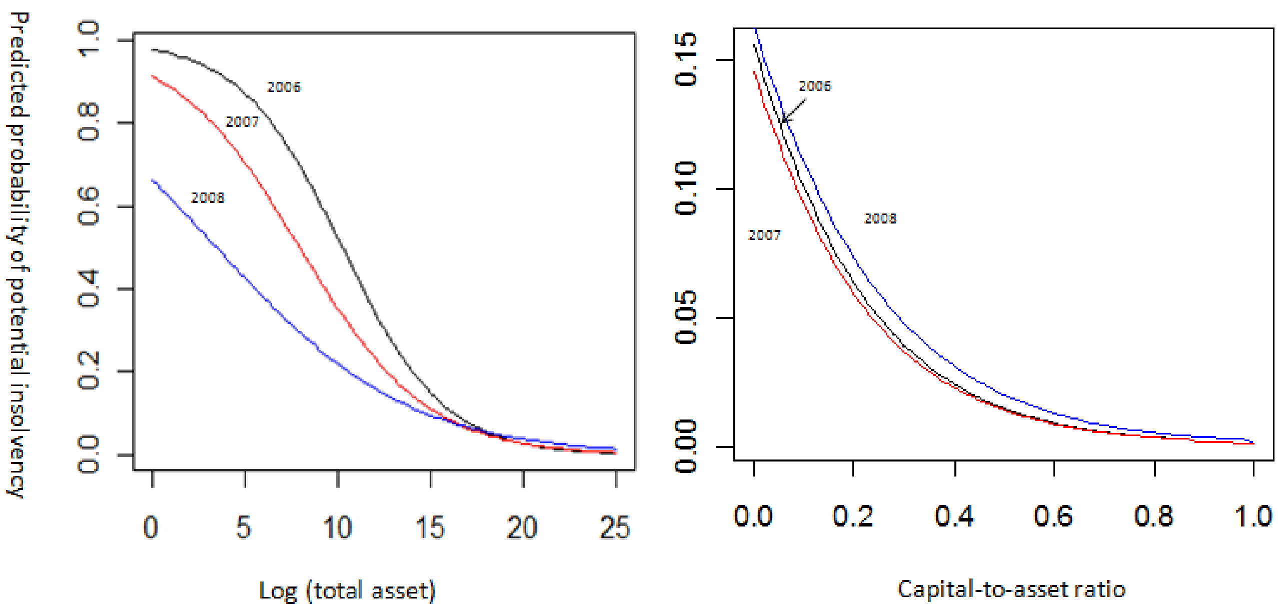

Figure 2: Predicted probability of potential insolvency of life insurer versus insurer's size (logarithm of total assets) and capital-to-asset ratio.

- *Risk measures*

Annuity product risk and health product risk work differently when affecting a life insurer's potential insolvent rate. All estimates of the regression coefficient for health product risk are significant from zeros while none of the estimates of regression coefficient for annuity product risk are significantly different from zero during the 2006-2008 period. This finding supports the opinion stated by Baranoff and Sager (2002) that health products pose greater risks than other products that are sold by life insurers. In fact, in the RBC ratio calculation, health writings receive a high penalty weight while annuity writings receive zero weight. The estimates of the regression coefficient for health product risk vary from 2.1139 to 3.0233, and the corresponding odds estimates vary from 8.28 to 20.538. That is, for every 0.01 unit increase in health product risk, the logits of potentially insolvent rates are expected to increase by 0.0211 to 0.0302 units on average, or the odds of potentially insolvent rates are expected to increase by a multiplicative factor of 1.021 to 1.031, holding all other explanatory variables constant during the 2006 to 2008 period.

Regarding asset risks, the estimate of the regression coefficient for regulatory asset risks is significantly different from zero only in year 2006, while none of the other estimates of the regression coefficient for regulatory asset risk and opportunity asset risks during the 2006-2008 period is significantly different from zero. This above finding suggests that asset risks might not play an important role in determining the insurer's potential insolvent rate.

All p-values of the regression coefficient for the proportion of reinsurance writings out of total writings are considerably less than 0.05. This indicates that there is a significant relationship between the insurer's potentially insolvent rate and its proportion of reinsurance writings. The estimates of the regression coefficient vary from 1.4541 to

2.533, and odds estimates vary from 4.281 to 12.591. That is, for every 0.01 unit increase in proportion of reinsurance writings, the logits of potentially insolvent rates are expected to increase by 0.0145 to 0.0253 units on average, or the odds of potentially insolvent rates are expected to increase by a multiplicative factor of 1.015 to 1.026, holding all other explanatory variables constant during the 2006 to 2008 period.

The predicted probabilities of potential insolvency versus health product risk and the proportion of products reinsured are plotted for three years as shown in Figure 2. Other explanatory variables are held constant at mean values for continuous variables or at 1 for indicator dummy variables (i.e. member of affiliated group and stock company).

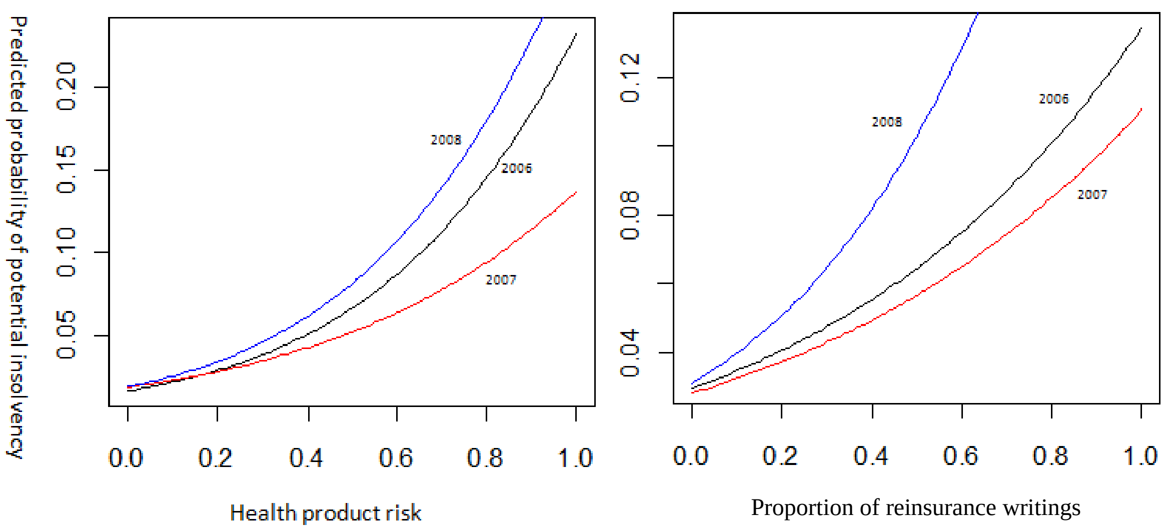

Figure 3: Predicted probability of potential insolvency of life insurer versus health product risk and proportion of reinsurance writings.

3.3 LOGISTIC MODEL ASSUMPTIONS

The logistic regression model, unlike the linear regression model, does not have restrictive assumptions such as a linear relationship between response and explanatory variables; normality of response variable and residual distributions; and homoscedasticity of residuals. However, some assumptions still apply, such as independency between records and absence of high multicollinearity.

- *Independency*

Two potential correlation problems exist in the NAIC database. The first one is that the response variable, which is the potentially insolvent indicator based on the RBC ratio, is an autocorrelated time series. Correlation is expected within the same company for successive years since the RBC ratio is likely to be similar and dependent on the value of the preceding year within the same company. To avoid the time series, the analysis is performed year by year separately, and the results are compared afterwards.

The second problem is a cross-sectional correlation between members of the same group of affiliated companies. These companies are likely to behave in similar ways and correlation may exist between them in the same year. To deal with this correlation problem, an additional categorical variable Group, representing different affiliated groups, is introduced to the model, and the effect of group dummy variables as additional predictors is examined for the years 2006, 2007 and 2008, separately. This is actually a test of whether each group should have its own intercept in the model. The p-values for tests on the global null hypothesis that betas are zeros and the type 3 analysis of effect for Group as an additional variable are both much greater than 0.05. This suggests that Group is not a significant predictor, and the correlation between members of a group of affiliated companies is not manifest through additive shifts in the model equation.

- *Multicollinearity*

Multicollinearity in a logistic regression model is defined as excessive correlations between explanatory variables. The existence of multicollinearity inflates the variances of the estimates of the regression coefficients and results in a higher probability of false negatives when testing the coefficients. It could also result in an incorrect sign and magnitudes of estimates of regression coefficients and lead to incorrect models. There are a total of ten explanatory variables in this logistic regression model, and multicolliearity could be a potential concern. In order to detect multicollinearity, multicollineartiy diagnostic statistics produced using the linear regression procedure in SAS are examined and shown in Table 3.

Table 3. Multicollinearity diagnostic statistics.

Variable	Tolerance			Variance Inflation Factor		
	2006	2007	2008	2006	2007	2008
LogATotal	0.44284	0.41096	0.41457	2.25814	2.43331	2.41215
CAP	0.49935	0.48669	0.49616	2.00258	2.05471	2.0155
RetOnCap	0.99436	0.9847	0.93298	1.00567	1.01554	1.07184
ProdARisk	0.60829	0.60932	0.59034	1.64394	1.64117	1.69393
ProdHRisk	0.7361	0.69461	0.70475	1.35852	1.43966	1.41895
Preinsur	0.80473	0.76898	0.77913	1.24265	1.30043	1.28348
logpregarisk	0.63821	0.48163	0.47785	1.56688	2.0763	2.0927
logpopparisk	0.57437	0.40489	0.42858	1.74104	2.46978	2.3333
Ngroup	0.81258	0.80815	0.80428	1.23065	1.23739	1.24335
Ntype	0.94595	0.94975	0.94359	1.05714	1.05291	1.05978

The Variance Inflation Factor (VIF) is the number of times the variances of the corresponding estimates of the regression coefficients are increased due to multicollinearity as compared to the number of times with no multicollinearity. A common understanding is that harmful multicollinearity is detected if VIF exceeds 10 in the linear regression model. However, for weak models such as the logistic regression model, Allison (1999) suggested that a VIF exceeding 2.5 could indicate a potential harmful multicollinearity. Table 3 shows that the largest VIF for explanatory variables during the 2006-2008 period is 2.47, suggesting there is no harmful multicollinearity in this logistic regression model.

3.4 SUMMARY AND CONCLUSION

In this study, the logistic regression model is used to explore the relationship between the potentially insolvent status of a life insurer and a set of explanatory variables involving the insurer's size, capital, governance structure, membership of group of affiliated companies, and various risk measures during the 2006-2008 period.

The results suggest that the probability of insolvency for an individual insurer is significantly affected by its size, capital-to-asset ratio, returns on capital, health product risk and proportion of reisnurance writings. It could be also possibly affected by the insurer's regulatory asset risk. However, the results indicate that the probability is not significantly related to the insurer's annuity product risk, opportunity asset risk, governance structure and its membership in a group of affiliated companies.

The assumptions of independency and absence of multicolliearity are both valid for this logistic model, suggesting that the model is adequate and the conclusion is convincing. For demonstration purpose, a RBC ratio of 200 is used as a cut-off value to identify a "potentially insolvent" insurer, which could lead to a discrepancy between the predicted status and real status. However, this model could still be useful in identify the significant factors which affect an insurer's potentially insolvent status.

3.5 REFERENCES

Allison, P.D. (1999), Logistic Regression Using the SAS System: Theory and Application, Cary, NC, SAS Institute Inc.

Baranoff, Etti G. and Sager,Thomas W., 2002, The relations among asset risk, product risk, and capital in the life insurance industry, Journal of Banking & Finance 26 (2002) 1181–1197.

Freese, J., and Long, J. S. 2006. Regression Models for Categorical Dependent Variables Using Stata. College Station: Stata Press.

Long, J. S. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks: Sage Publications.

Appendix

SAS Syntax for logistic regression analysis used in this study:

```
ods html;
ods graphics on;

libname ins '\\austin.utexas.edu\disk\xx562\wtsprofile\Desktop';

data record;
set ins.y93_08;
if rbcratio<0 or rbcratio>10000 then delete;
if rbcratio<200 then rbc200=0;else rbc200=1;
if year>2006 then delete;
if year<2006 then delete;
logpregarisk=log(pregarisk);
logpopparisk=log(popparisk);
run;

proc logistic data=record;
model rbc200=logatotal cap retoncap prodarisk prodhrisk preinsur
logpregarisk logpopparisk ngroup ntype /rsq;
run;

proc reg data=record;
model rbc200=logatotal cap retoncap prodarisk prodhrisk preinsur
logpregarisk logpopparisk ngroup ntype /tol vif;
run;


ods graphics off;
ods html close;
```

Bibliography

Agresti, A. 2002. Categorical Data Analysis. New York: Wiley-Interscience.

Allison, P. D. (1999), Logistic Regression Using the SAS System: Theory and Application, Cary, NC, SAS Institute Inc.

Baranoff, E. G., 2004, Risk Management and Insurance. John Wiley & Sons, Inc.

Baranoff, E. G. and Sager, T. W., 2002, The relations among asset risk, product risk, and capital in the life insurance industry, Journal of Banking & Finance 26 (2002) 1181–1197.

Browne, M. J., Carson, J. M., Hoyt, R. E., 1999, Economic and Market Predictors of insolvencies in the Life-Health Insurance Industry. Journal of Risk and Insurance, Vol. 66, No. 4, pp. 643-659.

Cox, D. R., and Snell, E. J. 1989. Analysis of Binary Data. Second edition. London: Chapman & Hall.

Cummins, J. D., Grace, M. F., and Phillips, R. D., 1997, "Regulatory Solvency Prediction in Property-Liability Insurance: Risk-Based Capital, Audit Ratios, and Cash Flow Simulation." Center for Risk Management and Insurance Research Working Paper 97-4, Georgia State University.

Cummins, J. D., Scott E. H., and Robert W. K., 1995, "Insolvency Experience, Risk-Based Capital, and Prompt Corrective Action in Property-Liability Insurance." Journal of Banking and Finance, 19: 511-527.

Freese, J., and Long, J. S. 2006. Regression Models for Categorical Dependent Variables Using Stata. College Station: Stata Press.

Grace, M. F., Scott E. H., and Robert W. K., 1998, "Risk-Based Capital and Solvency Screening in Property-Liability Insurance: Hypotheses and Empirical Tests." Journal of Risk and Insurance, 65: 213-243.

Grace, M. F., Timme, S. G., 1992. An examination of cost economics in the US life insurance industry. Journal of Risk and Insurance 59, 72–103.

Hilbe, J. M. 2009. Logistic Regression Models. Chapman & Hall/CRC Press

Hosmer, D. W., and Lemeshow, S. 2000. Applied Logistic Regression. 2nd ed. John Wiley & Sons Inc.

Long, J. S. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks: Sage Publications.

NAIC, 2009, NAIC Capital Adequacy Task Force - Risk Based Capital Overview. www.naic.org/documents/committees_e_capad_RBCoverview.pdf

Pottier, S.W., Sommer, D.W., 1997. Life insurer risk-based capital measures: Does the source matter? Journal of Insurance Regulation 16 (2), 179–196.

Pottier, S.W., and Sommer, D.W., 2002, The Effectiveness of Public and Private Sector Summary Risk Measures in Predicting Insurer Insolvencies, Journal of Financial Services Research, 21(1): 101–116.

Powers, A. P., and Xie, Y. 1999. Statistical Methods for Categorical Data Analysis. Academic Press, Inc.

Ryan Jr., H.E., Schellhorn, C.D., 2000. Life insurer cost efficiency before and after implementation of the NAIC risk-based capital standards. Journal of Insurance Regulation 18 (3), 362–384.

Shrieves, R.E., Dahl, D., 1992. The relationship between risk and capital in commercial banks. Journal of Banking and Finance 16, 439–457.